Foundations of Modern Networking

SDN, NFV, QoE, IoT, and Cloud

By: William Stallings

Chapter 10

Quality of Service

Background

- Historically the Internet and other IP-based networks provided a best effort delivery service
  - This means that the network attempts to allocate its resources with equal availability and priority to all traffic flows, with no regard for application priorities, traffic patterns and load, and customer requirements
- To protect the network from congestion collapse and to guarantee that some flows do not crowd out other flows, congestion control mechanisms were introduced, which tended to throttle traffic that consumed excessive resources
  - One of the most important congestion control techniques is the TCP congestion control mechanism
- For each TCP connection between two end systems across a network, in each direction, a concept known as sliding window is used

Background

- Although TCP congestion control and other network congestion control techniques can reduce the risk of excessive congestion, these techniques do not directly address QoS requirements

- As the intensity and variety of traffic increased, various QoS mechanisms were developed, including:
  - Integrated Services Architecture (ISA)
  - Differentiated services (DiffServ), accompanied by service level agreements (SLAs)

- These mechanisms and services serve two purposes:

Allocate network resources efficiently so as to maximize effective capacity

Enable networks to offer different levels of QoS to customers on the basis of customer requirements

Background

- In a sophisticated environment the term best effort refers not to the network services as a whole but to a class of traffic treated in best effort fashion
  - All packets in the best effort traffic class are transmitted with no guarantee regarding the speed with which the packets will be transmitted to the recipient or that the data will even be delivered entirely
- Typically in a network that provides multiple levels of service, best effort is the classification for the lowest priority traffic
- For some applications a class of traffic known as lower than best effort, or lower effort (LE) may be used
  - LE classification permits a network operator to strictly limit the effect of LE traffic on best effort/normal or all other network traffic
  - This classification may be suitable for background data transfer applications or traffic that could be delayed to off-peak times

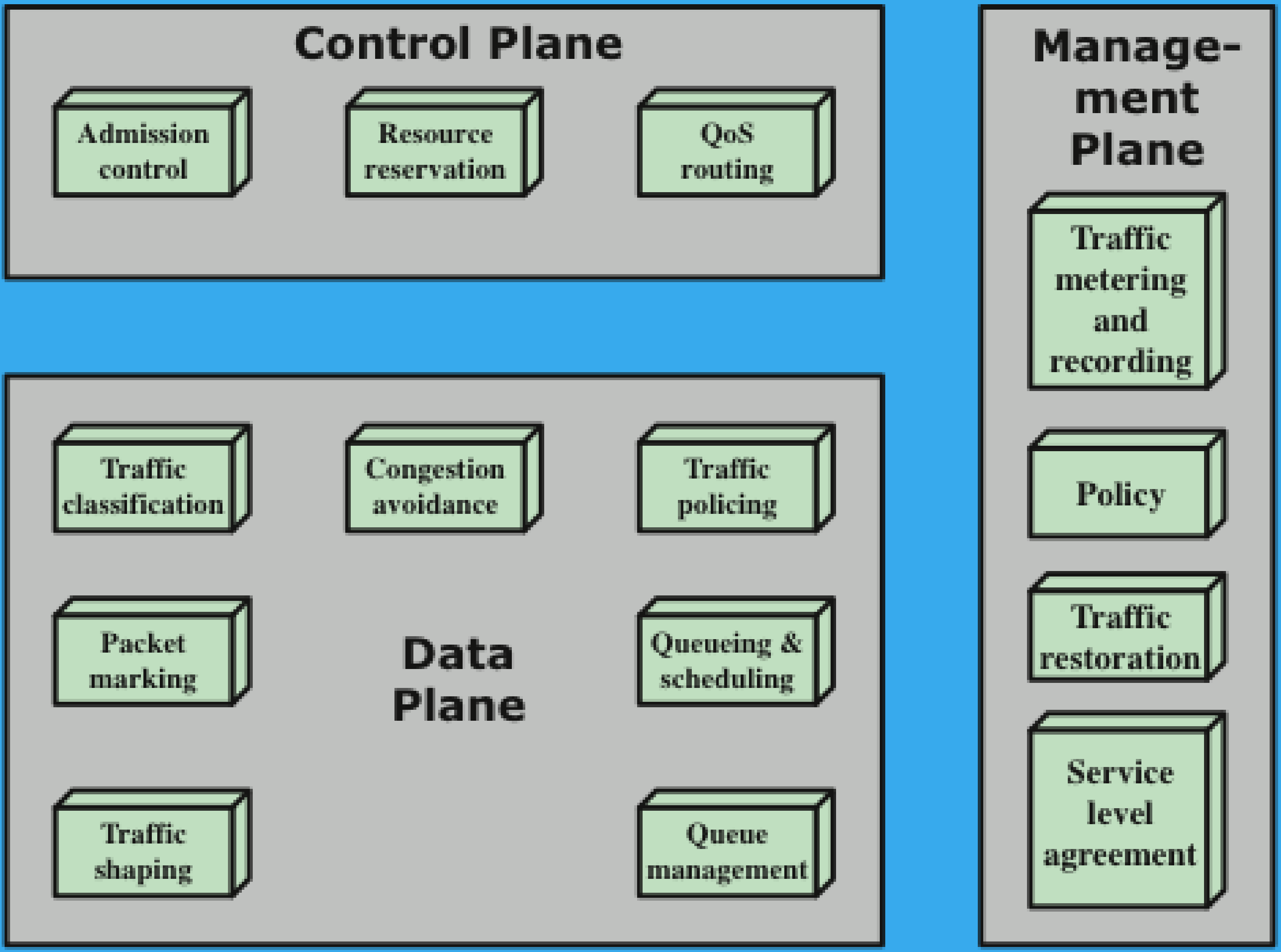


Figure 10.1 Architectural Framework for QoS Support

Control Plane

- The control plane is concerned with creating and managing the pathways through which user data flows
- It includes admission control, QoS routing, and resource reservation

Admission control

- Determines what user traffic may enter the network
- This may be in part determined by the QoS requirements of a data flow compared to the current resource commitment within the network

QoS routing

- Determines a network path that is likely to accommodate the requested QoS of a flow
- This contrasts with the philosophy of the traditional routing protocols, which generally are looking for a least-cost path through the network

Resource reservation

- A mechanism that reserves network resources on demand for delivering desired network performance to a requesting flow
- An example of a protocol that uses this capability is the Resource Reservation Protocol (RSVP)

Management Plane

- The management plane contains mechanisms that affect both control plane and data plane mechanisms
  - Deals with the operation, administration, and management aspects of the network
  - Includes SLAs, traffic restoration, traffic metering and recording, and policy

- Service level agreement (SLA)
  - Typically represents the agreement between a customer and a provider of a service that specifies the level of availability, serviceability, performance, operation, or other attributes of the service
- Traffic metering and recording
  - Concerns monitoring the dynamic properties of a traffic stream using performance metrics such as data rate and packet loss rate
  - Involves observing traffic characteristics at a given network point and collecting and storing the traffic information for analysis and further action; depending on the conformance level, a meter can invoke necessary treatment for the packet stream
- Traffic restoration
  - Refers to the network response to failures
  - This encompasses a number of protocol layers and techniques
- Policy
  - A category that refers to a set of rules for administering, managing, and controlling access to network resources
  - They can be specific to the needs of the service provider or reflect the agreement between the customer and service provider, which may include reliability and availability requirements over a period of time and other QoS requirements

Integrated Service Architecture (ISA)

- Purpose of ISA is to enable the provision of QoS support over IP-based internets
- Central design issue for ISA is how to share the available capacity in terms of congestion
- For an IP-based Internet that provides only a best effort service, the tools for controlling congestion and providing service are limited
  - In essence, routers have two mechanisms to work with:
    - Routing algorithm
    - Packet discard
- In ISA, each IP packet can be associated with a flow
  - A flow is a distinguishable stream or related IP packets that result from a single user activity and requires the same QoS

Integrated Service Architecture (ISA)

- ISA makes use of the following functions to manage congestion and provide QoS transport:

Admission control

- For QoS transport ISA requires that a reservation be made for a new flow
- If the routers collectively determine that there are insufficient resources to guarantee the requested QoS, the flow is not admitted
- The protocol RSVP is used to make reservations

Routing algorithm

- The routing decision may be based on a variety of QoS parameters, not just minimum delay

Queuing discipline

- A vital element of the ISA is an effective queuing policy that takes into account the differing requirements of different flows

Discard policy

- A discard policy determines which packets to drop when a buffer is full and new packets arrive
- A discard policy can be an important element in managing congestion and meeting QoS guarantee

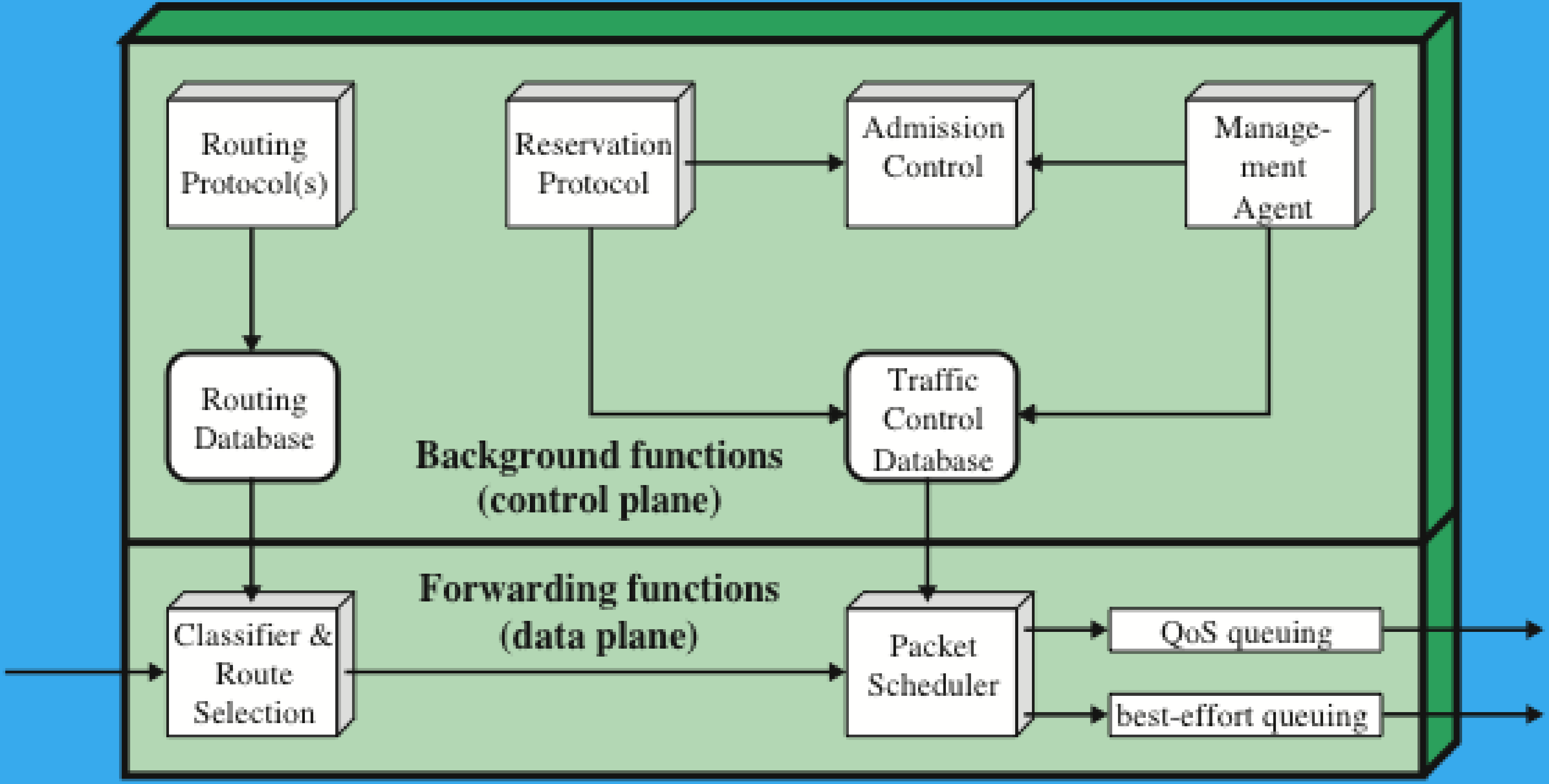
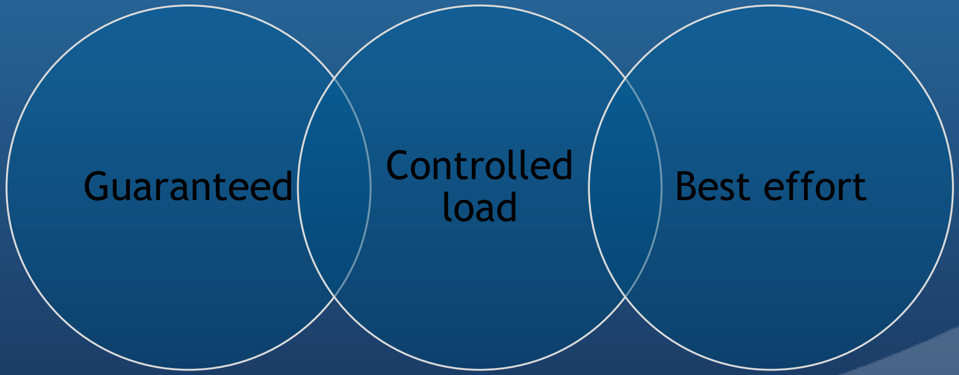


Figure 10.2 Integrated Services Architecture Implemented in Router

ISA Services

- ISA service for a flow of packets is defined on two levels
  - First, a number of general categories of service are provided, each of which provides a certain general type of service guarantees
  - Second, within each category, the service for a particular flow is specified by the values of certain parameters
- Together, these values are referred to as a traffic specification (Tspec)
- Three categories of service are defined:



Controlled Load

- The key elements of the controlled load service are:
  - The service tightly approximates the behavior visible to applications receiving best effort service under unloaded conditions
  - There is no specified upper bound on the queuing delay through the network
  - A very high percentage of transmitted packets will be successfully delivered
- The controlled load service guarantees that the network will set aside sufficient resources so that an application that receives this service will see a network that responds as if these real-time applications were not present and competing for resources
- Is useful for applications that have been referred to as adaptive real-time applications

Queuing Discipline

- An important component of an ISA implementation is the queuing discipline used at the routers
- The simplest approach that can be used by a router is a first-in, first-out (FIFO) queuing discipline at each output port
- A single queue is maintained at each output port
- When a new packet arrives and is routed to an output port, it is placed at the end of the queue
- As long as the queue is not empty, the router transmits packets from the queue, taking the oldest remaining packet next
- Drawbacks to the FIFO queuing discipline are:
  - No special treatment is given to packets from flows that are of higher priority or are more delay sensitive
  - If a number of smaller packets are queued behind a long packet, FIFO queuing results in a larger average delay per packet than if the shorter packets were transmitted before the longer packet
  - A selfish TCP connection, which ignores the TCP congestion control rules, can crowd out conforming connections

Queuing Discipline

- To overcome the drawbacks of FIFO queuing, a number of more complex routing algorithms have been implemented in routers
- These algorithms involve the use of multiple queues at each output port and some method of prioritizing the traffic to provide better service
- Typical of the networking industry are the routers from Cisco which, in addition to FIFO, include the following queuing approaches:
  - Priority queuing (PQ)
    - Each packet is assigned a priority level, and there is one queue for each priority level
  - Custom queuing (CQ)
    - Is designed to allow various applications or organizations to share the network among applications with specific minimum throughput or latency requirements
  - Flow-base weighted fair queuing (WFQ)
    - Creates flows based on a number of characteristics in a packet, including source and destination addresses, socket numbers, and session identifiers
  - Class-based weighted fair queuing (CBWFQ)
    - Allows a network administrator to create minimum guaranteed bandwidth classes

Differentiated Services (DiffServ)

- The differentiated services (DiffServ) architecture (RFC 2475) is designed to provide a simple, easy-to-implement, low-overhead tool to support a range of network services that are differentiated on the basis of performance
- Several key characteristics of DiffServ contribute to its efficiency and ease of deployment
    - IP packets are labeled for differing QoS treatment using the existing IPv4 or IPv6 DSField
    - A service level specification (SLS) is established between the service provider and the customer prior to the use of DiffServ
    - A traffic conditioning specification (TCS) is a part of the SLS that specifies traffic classifier rules and any corresponding traffic profiles and metering, marking, discarding/shaping rules which are to apply to the traffic stream
    - DiffServ provides a built-in aggregation mechanism; all traffic with the same DiffServ octet is treated the same by the network service
    - DiffServ is implemented in individual routers by queuing and forwarding packets based on the DiffServ octet; routers deal with each packet individually and do not have to save state information on packet flows
- DiffServ is the most widely accepted QoS mechanism in enterprise networks

Table 10.1 Terminology for Differentiated Services

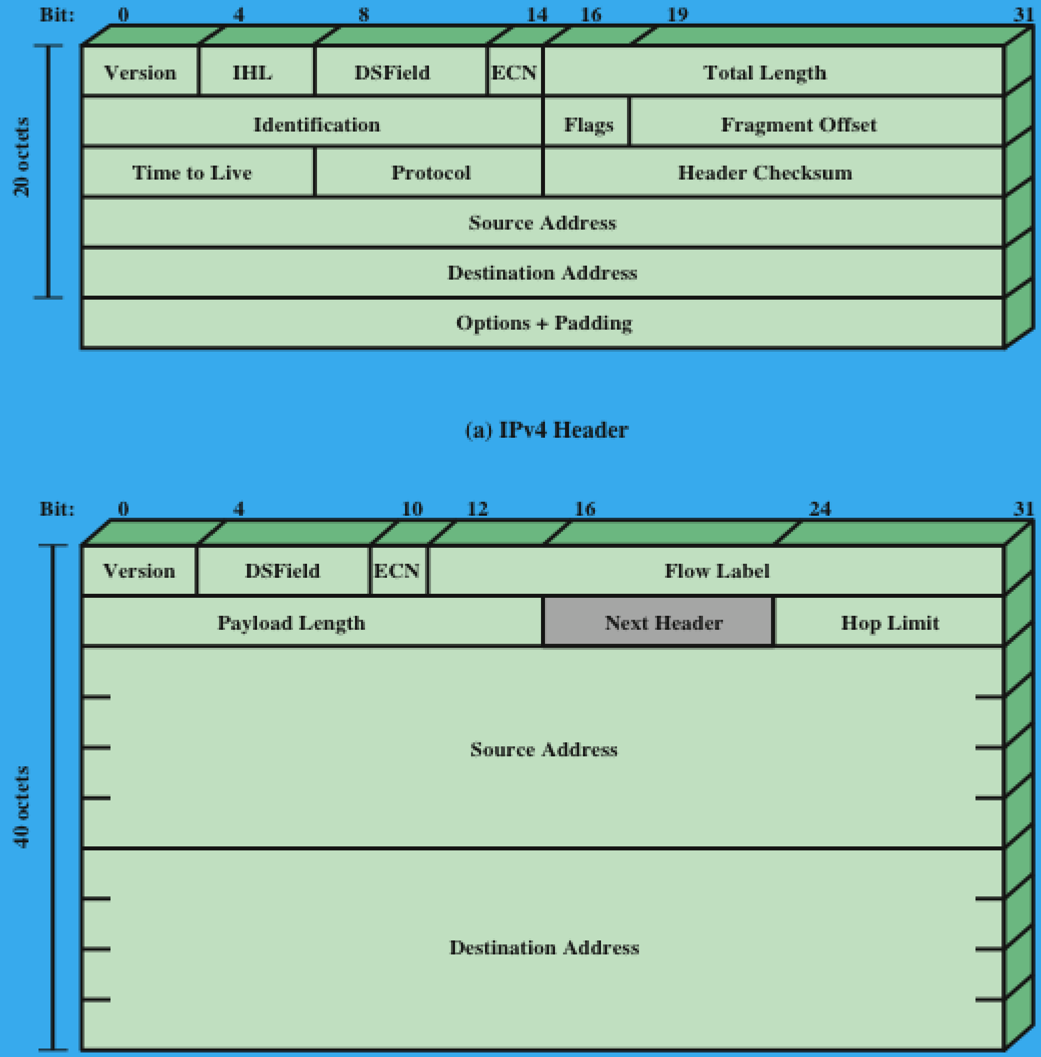
Behavior Aggregate	A set of packets with the same DS codepoint crossing a link in a particular direction.
Classifier	Selects packets based on the DS field (BA classifier) or on multiple fields within the packet header (MF classifier).
DS Boundary Node	A DS node that connects one DS domain to a node in another domain
DSField	The six most significant bits of the (former) IPV4 TOS octet or the (former) IPV6 Traffic Class octet.
DS Codepoint	A value that is encoded in the DSField.
DS Domain	A contiguous (connected) set of nodes, capable of implementing differentiated services, that operate with a common set of service provisioning policies and per-hop behavior definitions.
DS Interior Node	A DS node that is not a DS boundary node.
DS Node	A node that supports differentiated services. Typically, a DS node is a router. A host system that provides differentiated services for applications in the host is also a DS node.
Dropping	The process of discarding packets based on specified rules; also called **policing**.
Marking	The process of setting the DS codepoint in a packet. Packets may be marked on initiation and may be re-marked by an en route DS node.
Metering	The process of measuring the temporal properties (e.g., rate) of a packet stream selected by a classifier. The instantaneous state of that process may affect marking, shaping, and dropping functions.
Per-Hop Behavior (PHB)	The externally observable forwarding behavior applied at a node to a behavior aggregate.
Service Level Agreement (SLA)	A service contract between a customer and a service provider that specifies the forwarding service a customer should receive.
Shaping	The process of delaying packets within a packet stream to cause it to conform to some defined traffic profile.
Traffic Conditioning	Control functions performed to enforce rules specified in a TCA, including metering, marking, shaping, and dropping.
Traffic Conditioning Agreement (TCA)	An agreement specifying classifying rules and traffic conditioning rules that are to apply to packets selected by the classifier.

Services

- The DiffServ type of service is provided within a DiffServ domain, which is defined as a contiguous portion of the Internet over which a consistent set of DiffServ policies are administered

- The services provided across a DiffServ domain are defined in an SLA
  - Once the SLA is established, the customer submits packets with the DiffServ octet marked to indicate the packet class
  - The service provider must ensure that the customer gets at least the agreed QoS for each packet class

- If a customer submits packets intended for destination within the DIffServ domain, the DiffServ domain is expected to provide the agreed service
  - If the destination is beyond the customer's DiffServ domain, the DiffServ domain will attempt to forward the packets through other domains, requesting the most appropriate service to match the requested service



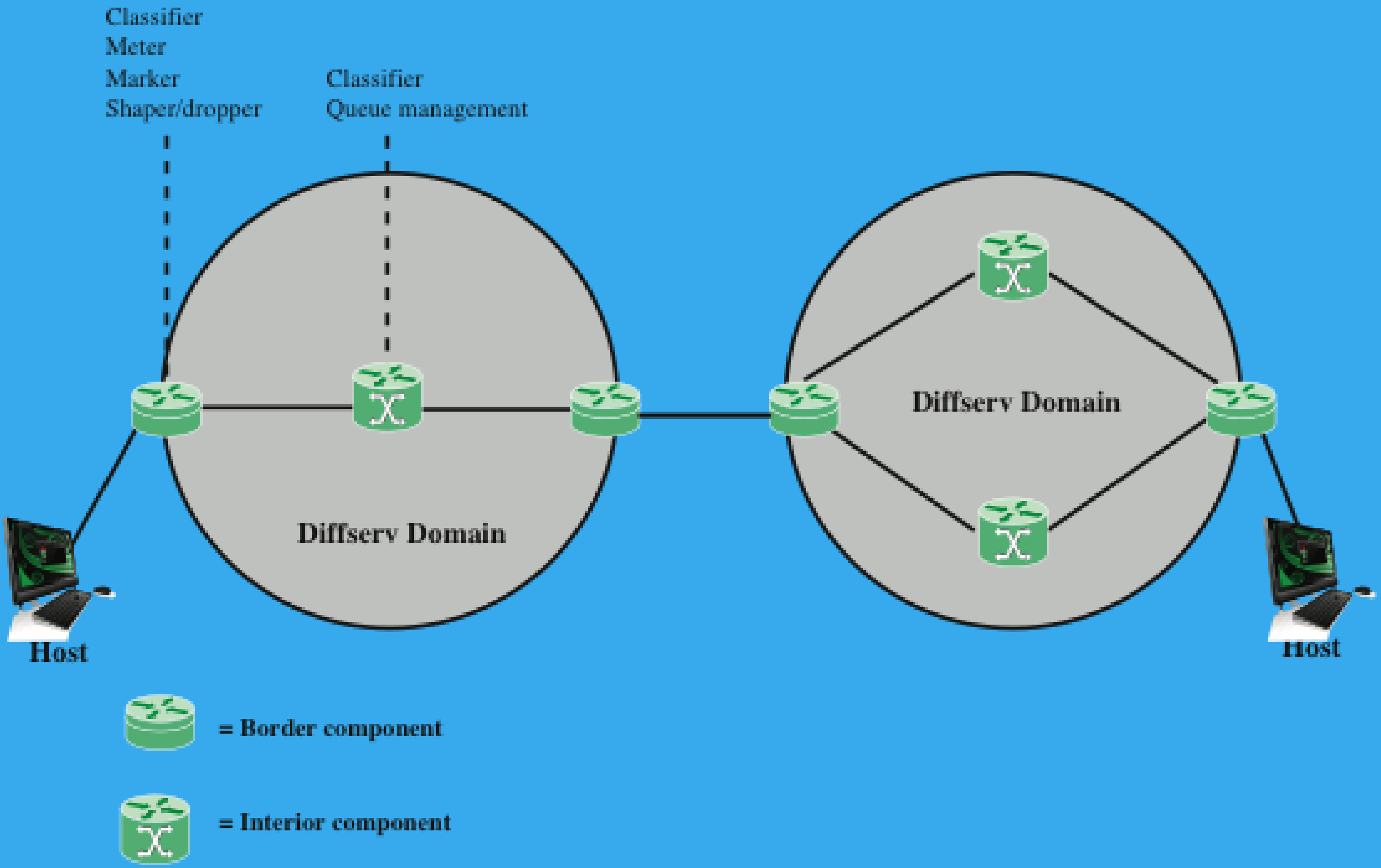
(a) IPv4 Header

(b) IPv6 Header

DSFeild = Differentiated services field
ECN = Explicit congestion notification field

Note: The 8-bit DSField/ECN fields were formerly known as the Type of Service field in the IPv4 header and the Traffic Class field in the IPv6 header.

Figure 10.3 IP Headers



Classifier
Meter
Marker
Shaper/dropper

Classifier
Queue management

Diffserv Domain

Diffserv Domain

Host

Host

= Border component

= Interior component

Figure 10.4 Diffserv Domains

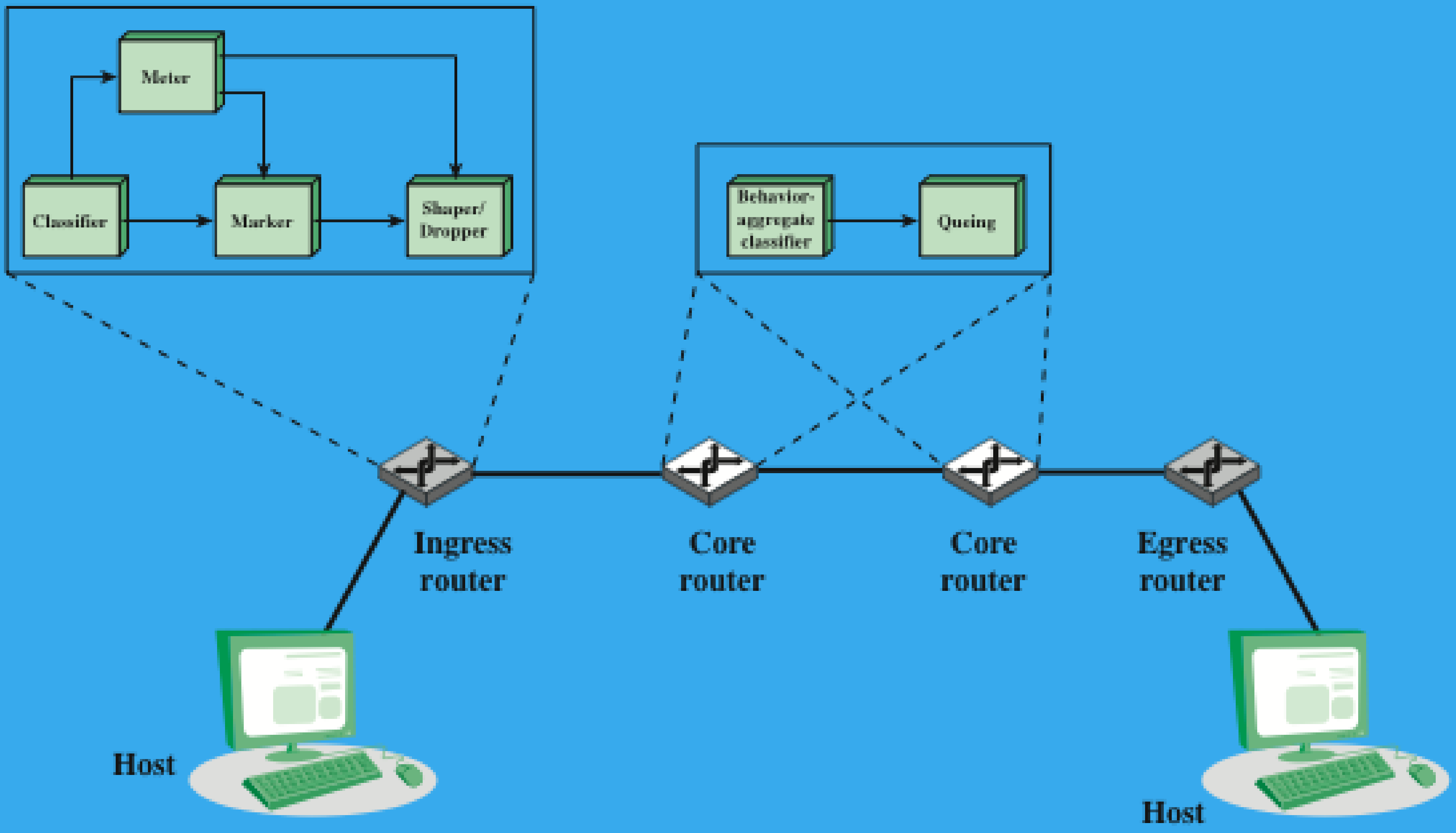
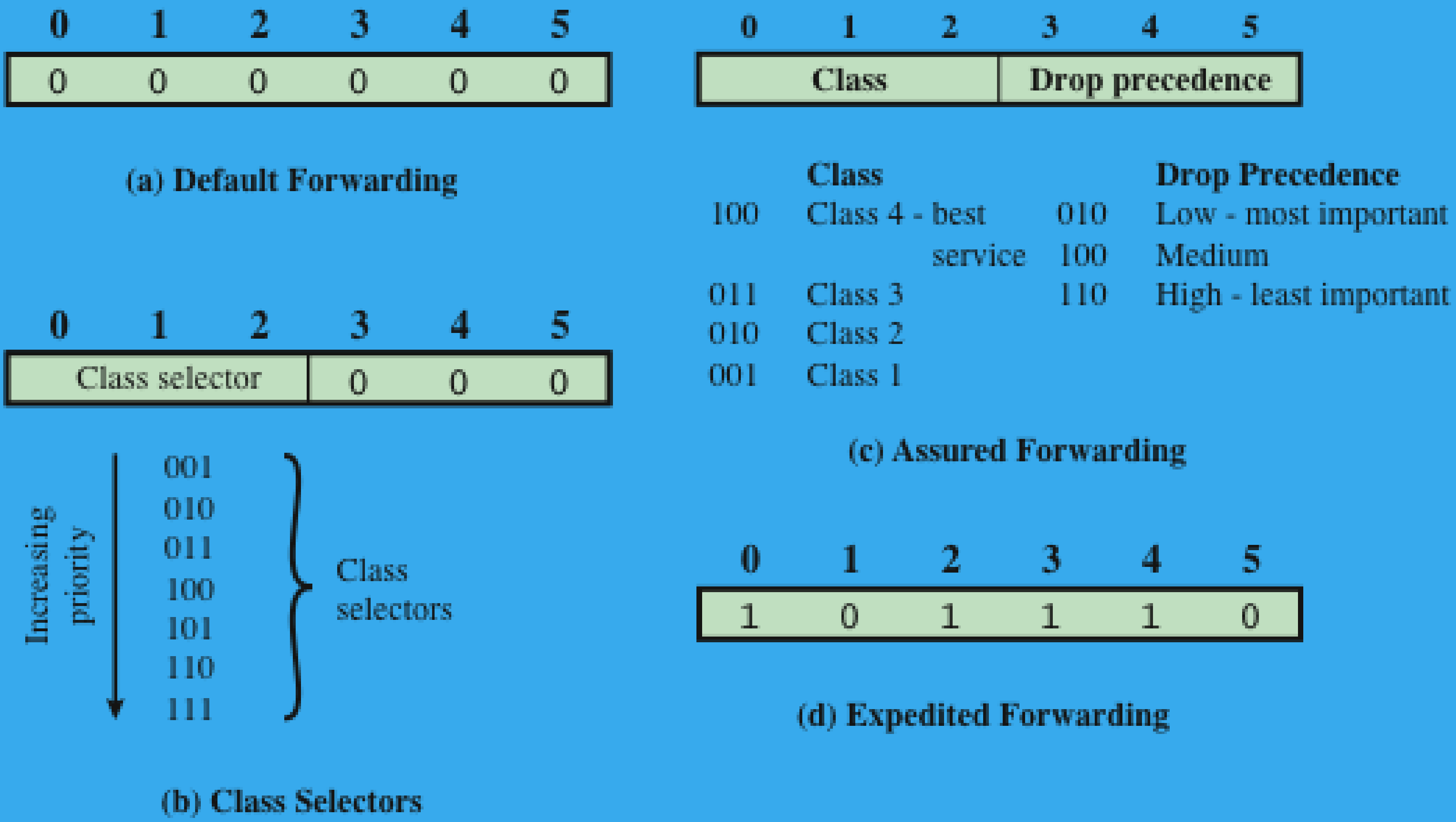


Figure 10.5 Diffserv Functions



0	1	2	3	4	5
0	0	0	0	0	0

(a) Default Forwarding

0	1	2	3	4	5
Class selector			0	0	0

(b) Class Selectors

0	1	2	3	4	5
Class			Drop precedence		

	Class		Drop Precedence
100	Class 4 - best service	010	Low - most important
		100	Medium
011	Class 3	110	High - least important
010	Class 2		
001	Class 1		

(c) Assured Forwarding

0	1	2	3	4	5
1	0	1	1	1	0

(d) Expedited Forwarding

Figure 10.6 Diffserv Forwarding Behavior Classes and Corresponding DSField Encoding

Default Forwarding PHB

- The default class, referred to as default forwarding (DF), is the best effort forwarding behavior in existing routers
- Such packets are forwarded in the order that they are received as soon as link capacity becomes available
- If other higher-priority packets in other DiffServ classes are available for transmission, the latter are given preference over best effort default packets
- Application traffic in the Internet that uses default forwarding is expected to be elastic in nature
- The sender of traffic is expected to adjust its transmission rate in response to changes in available rate, loss, or delay

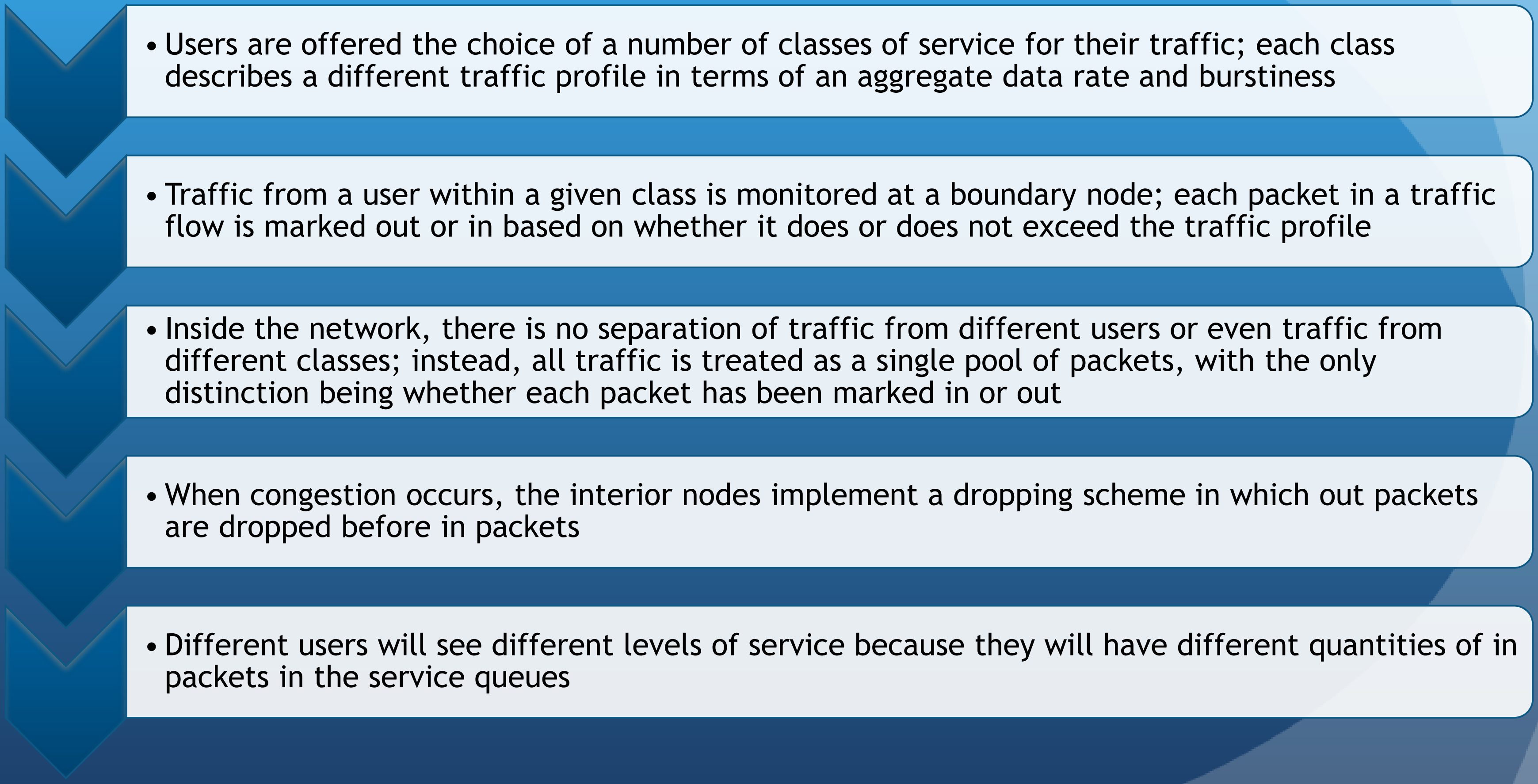
Expedited Forwarding PHB

- RFC 3246 defines the expedited forwarding (EF) PHB as a building block for low-loss, low-delay, and low-jitter end-to-end services through DiffServ domains
- Unless the internet is grossly oversized to eliminate all queuing effects, care must be taken in handling traffic for EF PHB to ensure that queuing effects do not result in loss, delay, or jitter above a given threshold
- RFC 3246 declares that the intent of the EF PHB is to provide a PHB in which suitable marked packets usually encounter short or empty queues
- The general concept outlined in RFC 3246 is:
  - The border nodes control the traffic aggregate to limit its characteristics to some predefined level; interior nodes must treat the incoming traffic in such a way that queuing effects do not appear

Assured Forwarding PHB

- The assured forwarding (AF) PHB is designed to provide a service superior to best effort but one that does not require the reservation of resources within an Internet and does not require the use of detailed discrimination among flows from different users

- The AF PHB is more complex than explicit allocation



- Users are offered the choice of a number of classes of service for their traffic; each class describes a different traffic profile in terms of an aggregate data rate and burstiness
- Traffic from a user within a given class is monitored at a boundary node; each packet in a traffic flow is marked out or in based on whether it does or does not exceed the traffic profile
- Inside the network, there is no separation of traffic from different users or even traffic from different classes; instead, all traffic is treated as a single pool of packets, with the only distinction being whether each packet has been marked in or out
- When congestion occurs, the interior nodes implement a dropping scheme in which out packets are dropped before in packets
- Different users will see different levels of service because they will have different quantities of in packets in the service queues

- The advantage of this approach is its simplicity

Class Selector PHB

- Codepoints of the form xxx000 are reserved to provide backward compatibility with the IPv4 precedence service
- The precedence field is set to indicate the degree of urgency or priority to be associated with a datagram
- If a router supports the precedence subfield, there are three approaches to responding:
  - Route selection
    - A particular route may be selected if the router has a smaller queue for that route or if the next hop on that route supports network precedence or priority
  - Network service
    - If the network on the next hop supports precedence, that service is invoked
  - Queuing discipline
    - A router may use precedence to affect how queues are handled

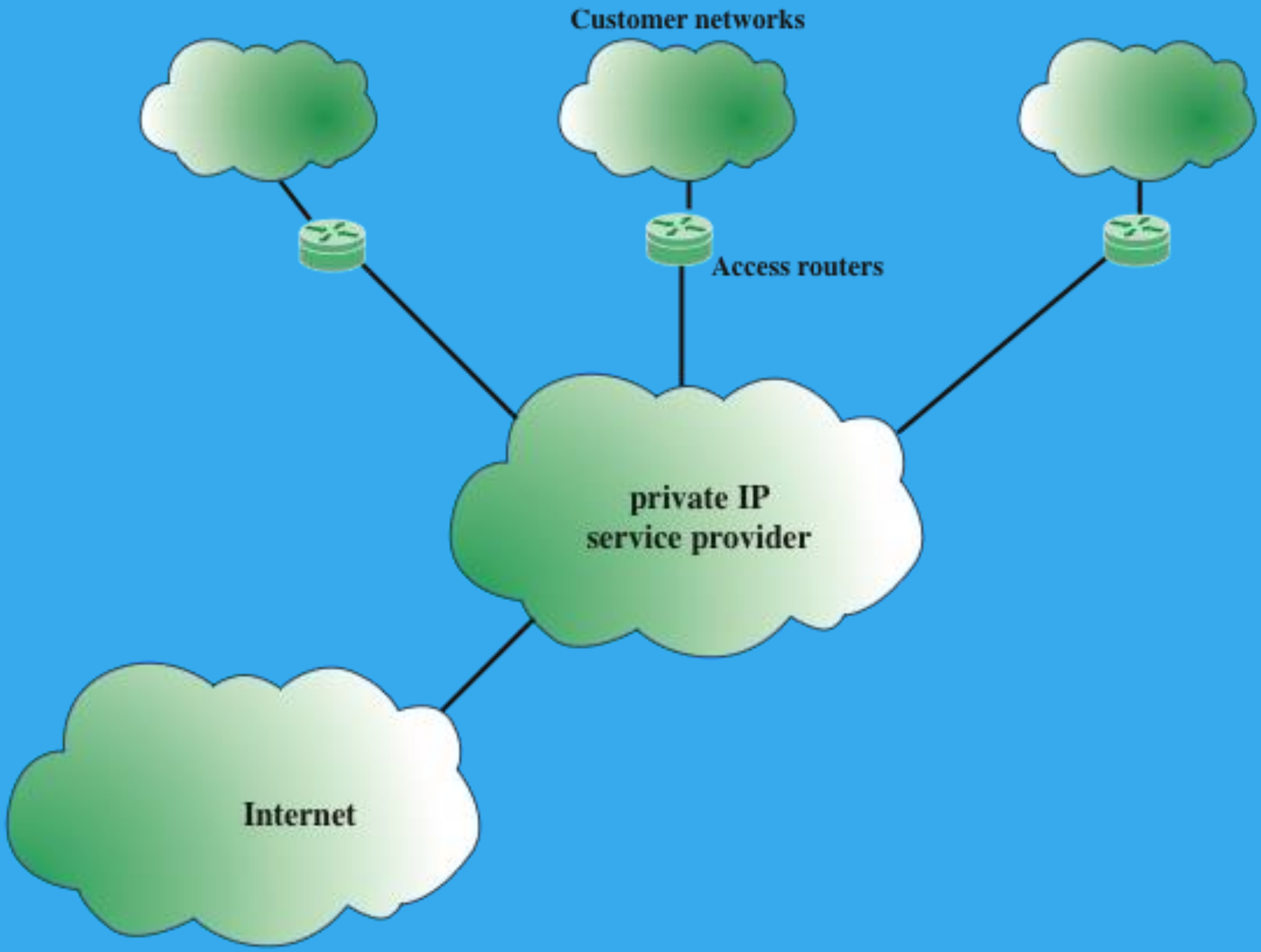
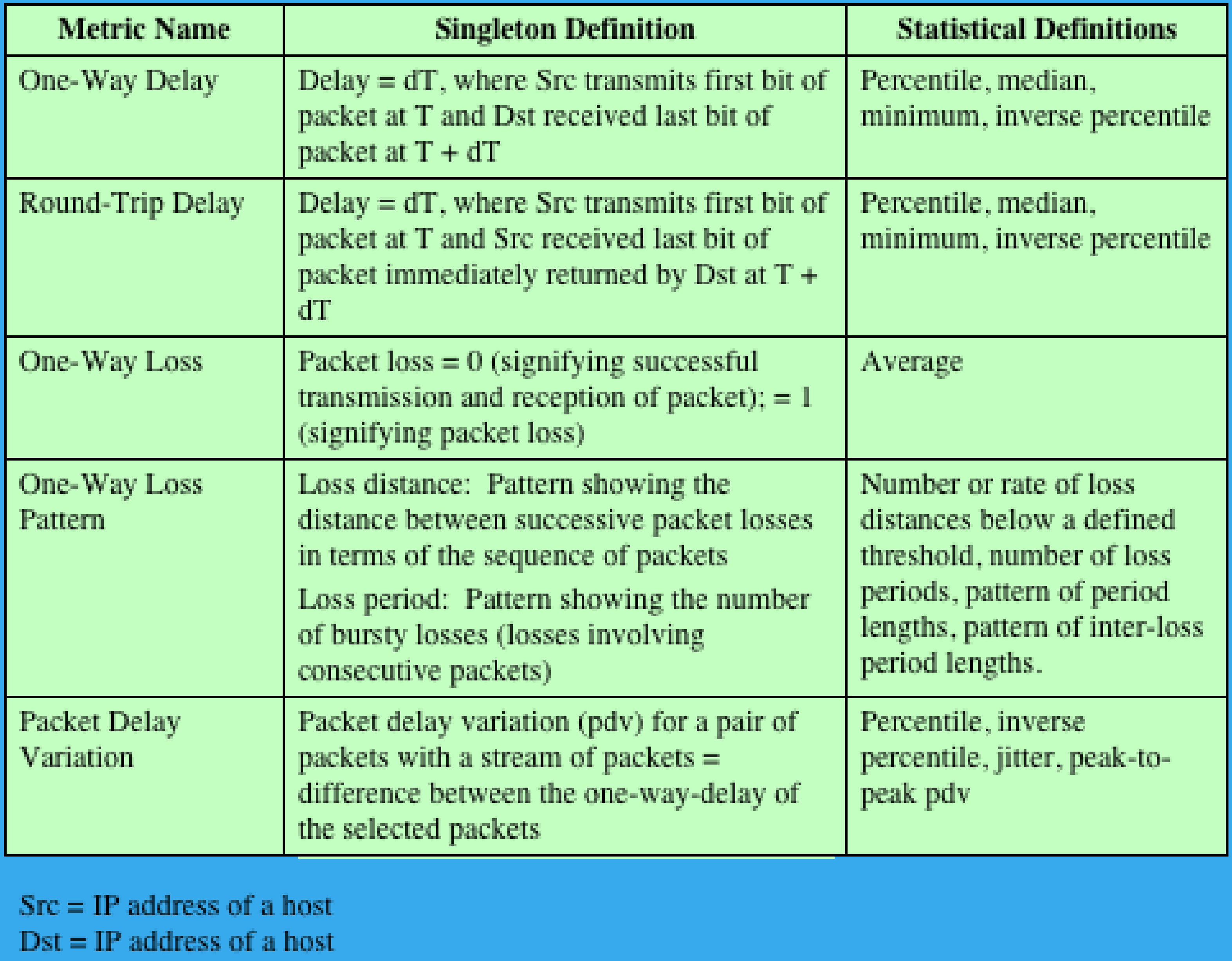


Figure 10.7 Typical Framework for Service Level Agreement

Table 10.2 IP Performance Metrics

(a) Sampled metrics

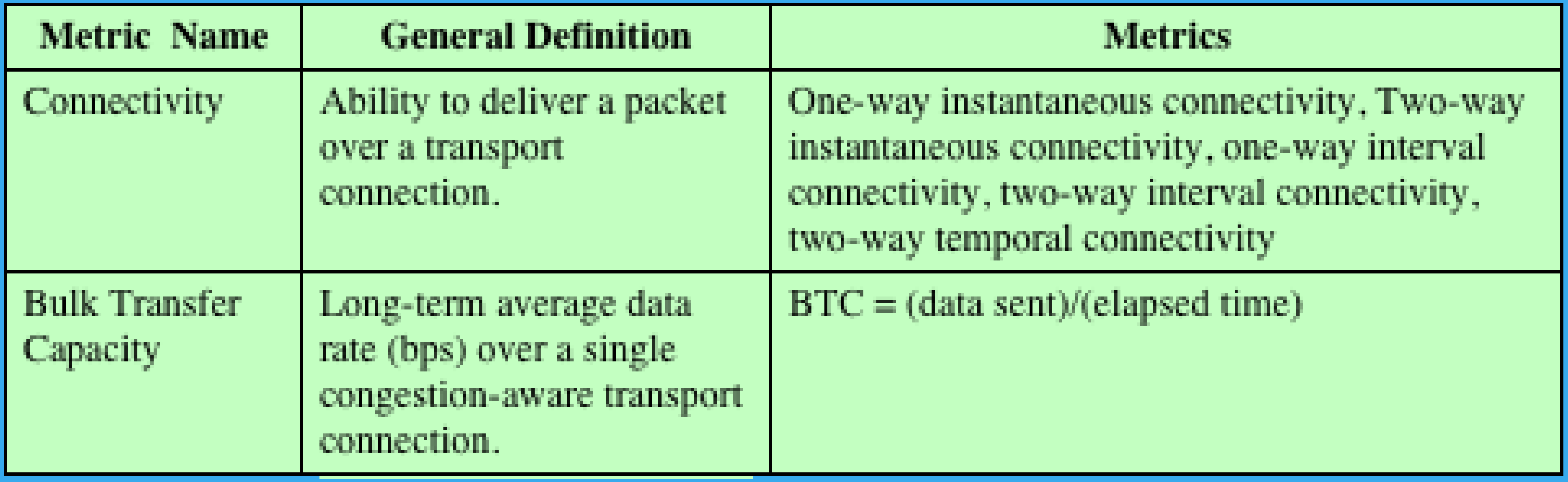


Metric Name	Singleton Definition	Statistical Definitions
One-Way Delay	Delay = dT, where Src transmits first bit of packet at T and Dst received last bit of packet at T + dT	Percentile, median, minimum, inverse percentile
Round-Trip Delay	Delay = dT, where Src transmits first bit of packet at T and Src received last bit of packet immediately returned by Dst at T + dT	Percentile, median, minimum, inverse percentile
One-Way Loss	Packet loss = 0 (signifying successful transmission and reception of packet); = 1 (signifying packet loss)	Average
One-Way Loss Pattern	Loss distance: Pattern showing the distance between successive packet losses in terms of the sequence of packets Loss period: Pattern showing the number of bursty losses (losses involving consecutive packets)	Number or rate of loss distances below a defined threshold, number of loss periods, pattern of period lengths, pattern of inter-loss period lengths.
Packet Delay Variation	Packet delay variation (pdv) for a pair of packets with a stream of packets = difference between the one-way-delay of the selected packets	Percentile, inverse percentile, jitter, peak-to-peak pdv

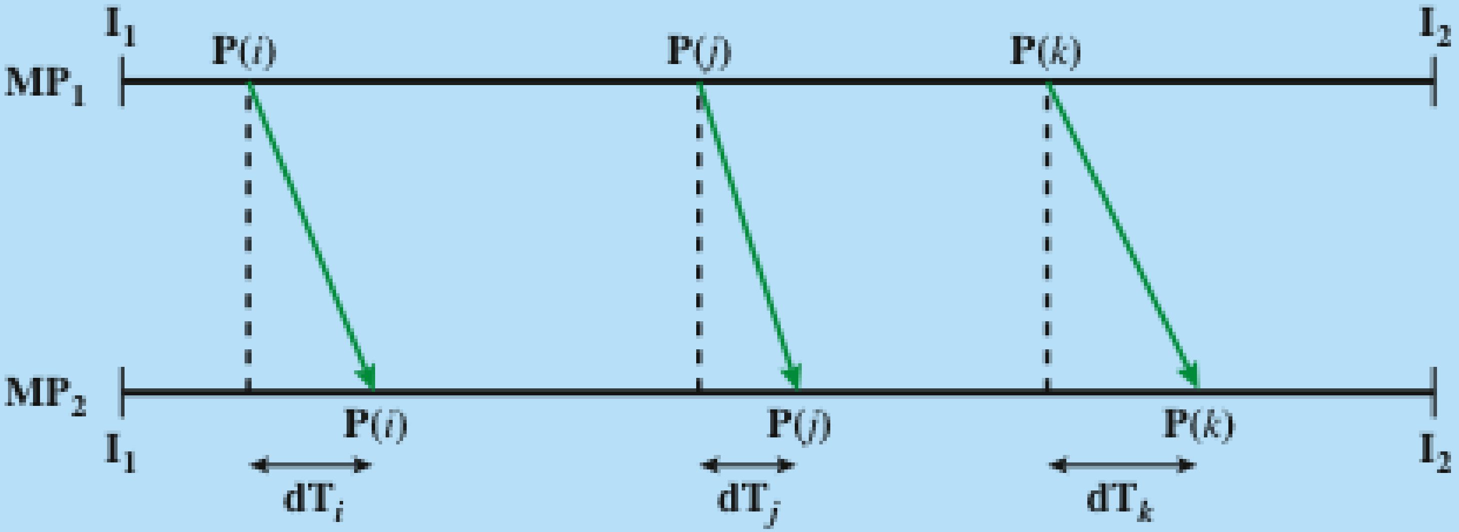
Src = IP address of a host
Dst = IP address of a host

Table 10.2 IP Performance Metrics

(b) Other metrics



Metric Name	General Definition	Metrics
Connectivity	Ability to deliver a packet over a transport connection.	One-way instantaneous connectivity, Two-way instantaneous connectivity, one-way interval connectivity, two-way interval connectivity, two-way temporal connectivity
Bulk Transfer Capacity	Long-term average data rate (bps) over a single congestion-aware transport connection.	BTC = (data sent)/(elapsed time)



I_1, I_2 = times that mark that beginning and ending of the interval in which the packet stream from which the singleton measurement is taken occurs.

MP_1, MP_2 = source and destination measurement points

$P(i)$ = ith measured packet in a stream of packets

dT_i = one-way delay for $P(i)$

Figure 10.8 Model for Defining Packet Delay Variation

OpenFlow QoS Support

- Queue Structures
  - An OpenFlow switch provides limited QoS support through a simple queuing mechanism
  - Queue configuration takes place outside the OpenFlow protocol, either through a command-line tool or through an external dedicated configuration protocol
- A data structure defines each queue
  - The data structure includes a unique identifier, port this queue is attached to, minimum data rate guaranteed, and maximum data rate
  - Counters associated with each queue capture the number of transmitted bytes and packets, number of packets dropped because of overrun, and the elapsed time the queue has been installed in the switch
  - The OpenFlow Set-Queue is determined beyond the scope of OpenFlow, thus any QoS feature must be implemented outside of OpenFlow

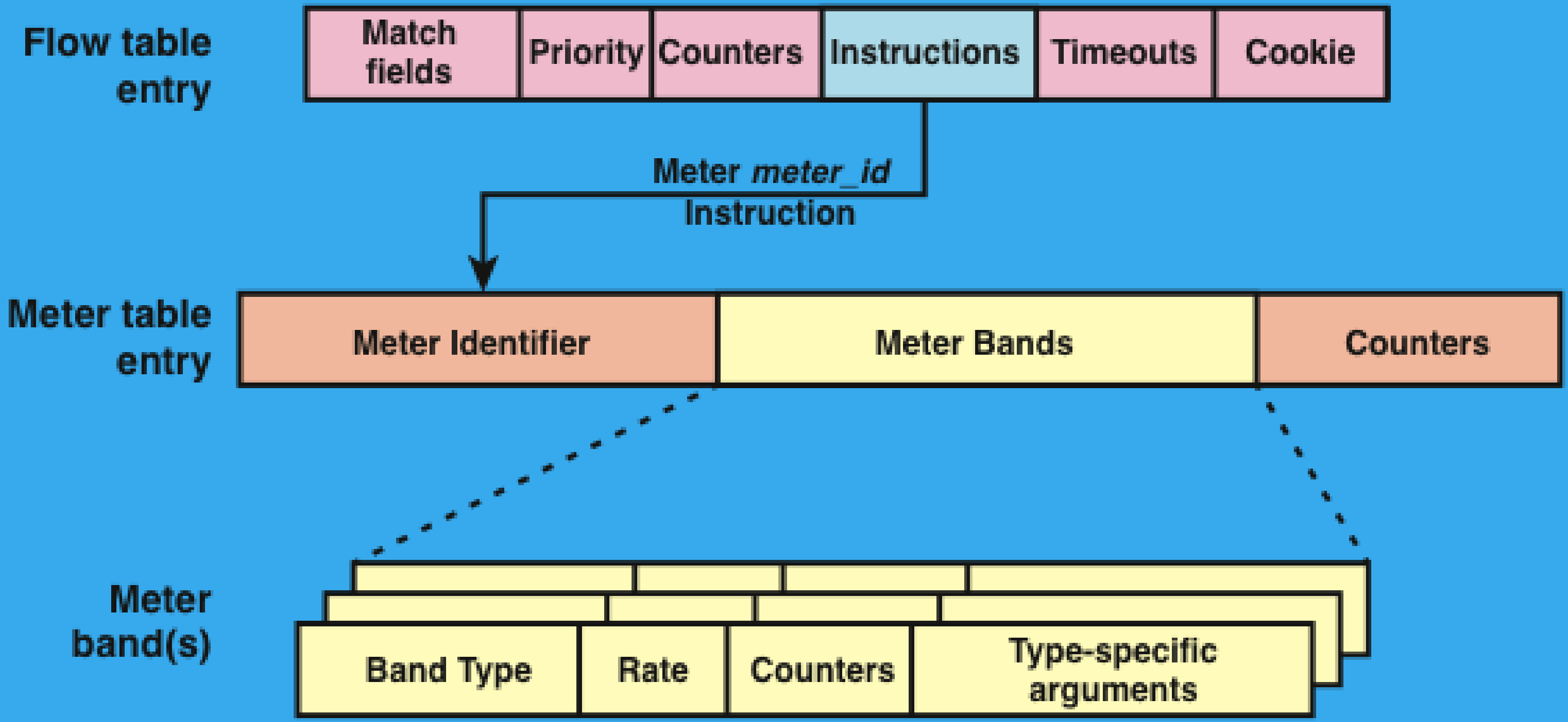


Figure 10.9 OpenFlow QoS-Related Formats

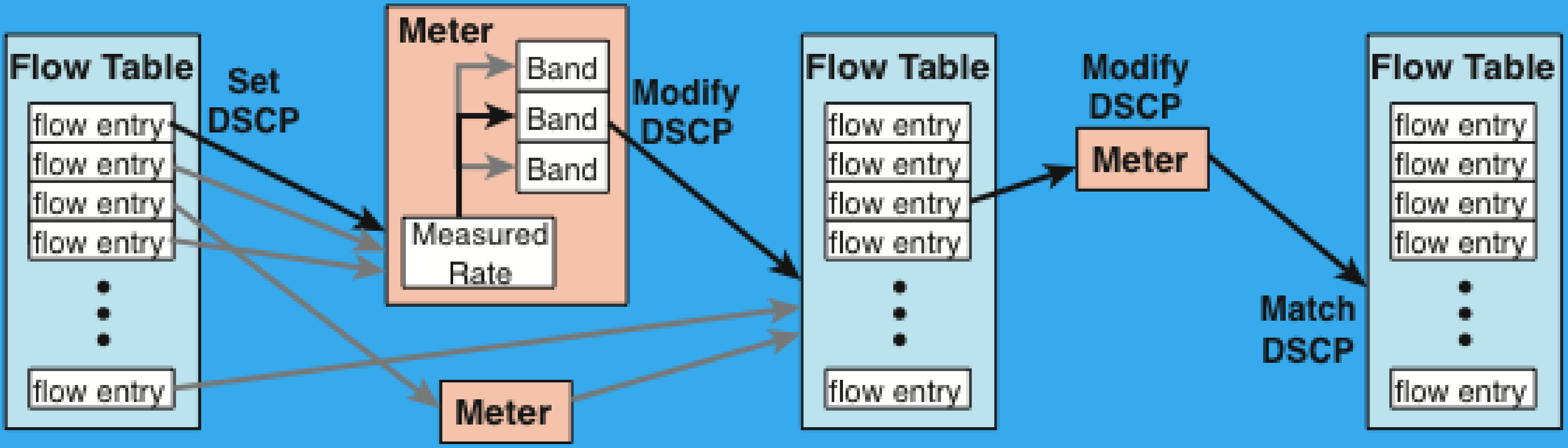


Figure 10.10 Diffserv Codepoint Metering

End of Chapter 10